**Fundamentals of Multimedia Computing**

**Chapter 1: Introduction to Multimedia Systems**

**Authors:**

**Gerald Friedland**

**and**

**Ramesh Jain**

**Draft for Comments**

**Chapter 1: Introduction to Multimedia Systems**

Only a few inventions in the history of civilization have impacted society in so many ways and at so many levelsas computers. Yet we have only seen the tip of the iceberg. ~~The nature of computing is rapidly evolving from simple alphanumeric data to rich multimedia experiences.~~ Where once we used computers for computing with simple alphanumeric data, we now use them primarily to exchange information and communicate. Computers are rapidly evolving as a means for gaining insights and sharing experiences across distance and time.

In the early days of computing, science fiction writersenvisaged computers as robots that would effectively use audio-visual data. Later, people dreamt of systems that could organize audio files, images, and video. Now people want to share perceptual experiences independent of time and distance. Within the next few years, most of the data stored on computers—at least with regard to storage size and bandwidth requirements—will be audio-visual. The fundamental medium for computing and communications will also be audio-visual.

Handling multimedia content requires incorporating concepts and techniques from various disciplines—from audio to visual processing, from communication theory to image databases, and from compression techniques to content analysis. Multimedia computing has consequently evolved as a collection of techniques from different disciplines.

Unfortunately, multimedia has become like the elephant in the fable about the elephant and the six blind men (see Figure 1). In this fable, each blind man has limited perspective due to some physical limitation. In real life, people impose limitations of perspective in many ways and hence—though naturally endowed with multiple sensory and cognitive faculties—functionally behave like these blind men. Like the blind men portrayed in the cartoon, each discipline perceives multimedia **in a limited aspect** **//from its own limited viewpoint//**. This has resulted in the field’s skewed development.

We use our five senses (sight, hearing, touch, smell, and taste) together with our abstract knowledge to form holistic experiences and extract information. Multimedia computing aims to develop communication techniques to allow holistic experiences from multiple sources and modalities of data and extract useful information in the context of various applications.

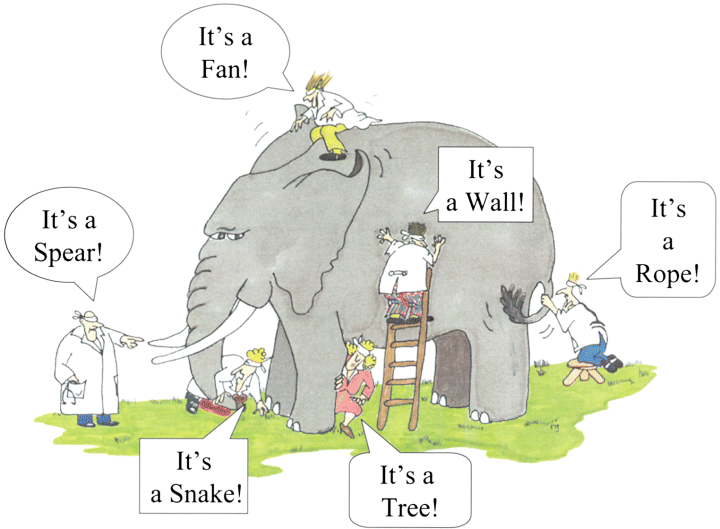


Figure 1: Multimedia is like an elephant. Looking at it from limited perspective leads to completely wrong characterizations.

This fragmented perspective of multimedia has slowed progress in understanding and effectively processing multimedia information, although the hardware used for processing it—ranging from sensors to bandwidth—has advanced rapidly. Multimedia computing should leverage correlated and contextual information from all sources to develop holistic and unified perspectives and experiences. It should focus on **//full multisensory//** experiences rather than partial experience, such as listening to a an audio-only sports commentary.

This book presents emerging techniques in multimedia computing from an experiential perspective in which each medium—audio, images, text, and so on—is a strong component of the complete exchange ofinformation or experience. Humans are the best functioning example of multimedia communication and computing—that is, we understand information and experiences through the unified perspective offered by our five senses. Our goal in this book is to present current techniques in computing and communication that will lead to the development of a unified and holistic approach to computing using heterogeneous data sources.

We identify the needs of emerging applications that use multifarious data sources of heterogeneous types. This book introduces the elements that are useful in defining and designing these emerging systems. It serves as an introduction to multimedia systems for engineers and researchers interested in understanding the elements of multimedia and their role in building specific applications.

# Communication in Human Society

The ability to communicate effectively is one of the main **factors** that distinguishes humans from animals and has been a major force in human evolution. Communication lets us share experiences and create, maintain, sustain, and propagate knowledge. As table 1 shows, human civilization has seen many influential inventions related to communicating experiences across space and time **//perhaps clarify the significance of space and time???//**.

|  |  |  |
| --- | --- | --- |
| **Invention** | **Resulting Application** | **Dimension** |
| Languages | Symbolic communication of experiences | Space |
| Written languages | Symbolic record of experiences | Time |
| Paper | Portability | Time and space |
| Print | Mass distribution | Time and space |
| Telegraph | Remote narrow communication | Space |
| Telephone | Remote analog communication | Space |
| Radio | Analog broadcasting of sound | Space |
| Television | Analog broadcasting of sight and sound | Space |
| Recording media | Analog recording | Time and space |
| Digital media | Machine enhancement and processing | Time and space |
| Internet | Personalized reception | Time and space |

Table 1: Communications-related inventions in human civilization.

Before people had language, they had no way to share experiences. At first, **//spoken//** language was just analog sounds from vocal cords. Eventually, language assumed a symbolic structure based on those sounds. People soon realized that experiences were important and should be stored for sharing with others. They thus invented written language as a system for representing sounds so other people could also share experiences. Cumbersome techniques such as stone tablets gave way to more practical storage devices and writing methods. Next came the development of paper and ink, and still more people began using the stored experiences that others had painstakingly recorded.

Then came one of the most influential inventions in our history—Gutenberg’s movable printing press. This invention enabled mass communication for the first time, and revolutionized society. Our current education system, our reliance on documents (such as newspapers) as a major source of communication, and libraries as public, government-supported institutions dedicated to storing knowledge, all stem from that one invention that appeared more than 500 years ago.

The telegraph, which allowed instantaneous communication of symbolic information over long distances, began to bring the world closer. This invention signaled the beginning of the global village. Telephones let us return to our natural communication medium—talking—while retaining all of the advantages of instantaneous remote communication. People could experience the emotions of the person on the other end of the connection—something symbol-based methods of writing and telegraph could only hint at.

Radio ushered in the wireless approach to sound and popularized sound as a medium for instantaneous mass communication. Television took communication a step further by appealing to our sense of sight as well as hearing. It was the first medium that let us experience with more than one sense and as such was able to more effectively key into our emotions. Video communication’s popularity is clearly due to its use of our two most powerful senses working in harmony to communicate experiences.

Storage and distribution technologies, such as magnetic tape, allowed the storage, preservation, and distribution of sound, again bringing us closer to natural experience. Video recording enhanced this experience significantly. Digital media further improved the quality of our experience. The Internet took information availability to a new dimension, providing us with experiential accounts of an unprecedented variety.

# Evolution of Computing and Communication Technology

The changes in the landscapes of both computing and communications have been overwhelming in the last few decades.

Just a few decades ago, a computing center was one of the most important buildings on a university or corporate campus. Access to this building, particularly to the holy room in which the computer functioned, was highly restricted. A computer occupied several rooms, if not floors, of a building, needed air conditioning, and required a specialized and trained staff to interact with it and run users’ programs on it. These computers cost millions of dollars. Figure 2 shows a popular computer from the late 1960s and early 1970s. Table 2 lists some of its important characteristics.



Figure 2: A 1960s-era computer.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Computer** | **Processing unit** | **Operating system** | **Core memory** | **Secondary memory** |
| 1960s-era computer | Could not do arithmetic, used look-up tables | No OS; human monitors controlled everything | 60 Kbytes | 2M characters |
| Modern handheld computer (iPhone) | ARM 620 MHz | iPhone OS | 128 Mbytes | 16 Gbytes |

Table 2: Comparison of early computers with those of a typical handheld computer.

Progress in processing, storage, networking, and software technology have changed computing beyond anyone’s expectations. Today, most people carry computers that are more powerful and sophisticated than the 1960s-era computer in their pockets. Figure 3 shows one such computer; Table 2 compares it to the early computer. Although this computer is more powerful and sophisticated than the one in Figure 2, it costs several thousandths of what the older version does, is easy to carry, and isn’t affected by climate **//correct?//**. Moreover, just about anyone can operate it, using it to solve their everyday computing and communications needs.



Figure 3: A handheld computer, similar to those most people carry.

Communications technology has experienced a similar overwhelming transformation. We’ve already discussed the historical perspective. Here, we focus on short-term technological improvements in one medium.

Consider the telephone. In its very early incarnations **//okay?//**, the telephone had limited use. Only a few people could afford to have one in their homes. Moreover, a house had one phone, and when you called someone you literally had to shout into the mouthpiece. During a long-distance call, latency made communication difficult. Either both parties spoke at the same time or each waited for the other, while an expensive meter ticked off minutes. People spent more time shouting “Hello! Hello!” than having a meaningful conversation. Now, users can talk on a phone while walking, running, driving, or flying in an airplane. Signal reception is so clear that you can whisper to a person on the other side of the globe. In addition, not only is your phone a voice communication device, but it is also your connection to a computer network, a camera, your calendar and address book, and soon, a video communication device.

# Driving Applications

To understand computing technology’s evolution to its current state, as well as to project its future evolution, consider the applications that have been and will be driving the technology’s development.

The first computer applications performed numerical computations using data in scientific applications, hence the name *computer*. Business was the next major driving application with so-called “data processing.” It brought alphanumeric processing and databases in focus for development. Major networking advances resulted in enterprise computing based on the traditional distributed processing approaches that eventually culminated in the Internet.

Personal computers were another major influence on computing. PCs ended reliance on a powerful central computer and put several applications, including word processing, spreadsheets, and electronic mail, in a completely new perspective. Combining personal computing and Internet connectivity led to one of the most amazing revolutions that human civilization has ever seen: the World Wide Web. The progress continued and laptop computers replaced most PCs. Laptops are now being replaced by **//an even more//** personal and sentient computing device—the mobile phone. These phones can be used for computing, communication, and much more. Moreover, they can use audio and visual mechanisms equally effectively as traditional alphanumeric computing. Slowly, they are being equipped with more diverse sensing mechanisms than humans have. These devices are true multimedia computing and communication devices.

# The Nature of Emerging Applications

Emerging computing and communication applications have clear differences from earlier applications. For example, in these emerging applications:

* Spatiotemporal and live data streams are the norm rather than the exception.
* A holistic picture of an event is more important than silos of isolated data.
* Users want insights and information that are independent of the medium and data source. That is, the medium is just the medium; the message is what’s important.
* Users do not want information that is not immediately relevant to their particular interests.
* Exploration, not query, is the predominant mode of interaction.

These applications are pushing computing to use primarily multimedia data from multiple sources. Moreover, these applications clearly demand that computing focus more on information, experiences, and understanding than on the medium or data source.

# Exploiting the Human-Machine Synergy

The evolving nature of data sources and desired operations can be captured in the matrix shown in Table 3. These relationships have profound implications for information and communication technologies (ICT). For example, databases are excellent for getting precise information from a single alphanumeric data destination. Visualization environments and interactive tools combined with data warehousing technology are useful in gaining insights from a precise alphanumeric source. In the last few years, search engines have made tremendous progress in finding information sources, particularly alphanumeric sources, inthe WWW environment, which is primarily an unstructured distributed information source. Going forward, most emerging applications will fall in the top-right quadrant: To gain insights from multiple heterogeneous sources, we need an experiential environment because it unites disparate data sources and frees decision-makers to explore their perceptions.

| Insight | Visualization | Experiental Environments |
| --- | --- | --- |
| **Information** | Databases | Search Engines |
|  | **Single Data Destination** | **Multiple Data Destination** |

Table 3: The changing nature of applications.

Current information environments actually work against the human-machine synergy. Humans are very efficient in conceptual and perceptual analysis but relatively weak in mathematical and logical analysis; computers are exactly the opposite. In an experiential environment, users *directly use their senses to observe data and information of interest related to an event, and they interact naturally with the data based on their particular set of interests in the context of that event.*

Experiential environments several important characteristics:

*They are direct.*An experiential environment provides a holistic picture of an event without using unfamiliar metaphors and commands. People are in a familiar environment and use natural actions based on commonly used operations and their anticipated results. In experiential environments, users easily and rapidly interpret the data presented and then interact with the dataset to get a modified dataset.

*They provide the**same query and presentation spaces.* Most current information systems use different query and presentation spaces. Popular search engines, for example, provide a box for entering keywords, and the system responds with a list of perhaps thousands of entries spanning hundreds of pages. A user has no idea how the entries on the first page relate to those on the 13th, how many times the same entry appears, or even how th*e entries on the same page relate to each other*. Contrast this to a spreadsheet. A user articulates a query by changing certain data that is displayed in the context of other data items. This action results in a new sheet showing new relationships. Here, query and presentation spaces are the same. These systems are called What-You-See-Is-What-You-Get (WYSIWYG).

*They consider* *both the user state and context.* Systems should know the user’s state and context and present information that is relevant to the user in that state and context.People operate best **//when they are//** in known contexts and **//they//** do not like instantaneous context switching. Information systems, including databases, should be scalable and efficient. These considerations led to the design of *stateless* systems, such as relational databases. However, this statelessness is why most Internet search engines are so dissatisfying. They don’t remember the user’s state. **//check changes//**

*They promote perceptual analysis and exploration.* Text-based systems provide abstract information in visual form. Experiential systems let user analyze, explore, and interact with their environment using all of their senses, and thus are more compelling and easier to understand.

Multimedia systems will play a key role in creating experiential environments in diverse applications. Currently, video games provide the best experiential environments. These games effectively use audio, video, and tactile media to create compelling interactive environments.

# Multimedia Computing Systems

Multimedia computing started gaining serious attention from researchers and practitioners in the 1990s. Before 1991, people talked about multimedia, but the computing power, storage, bandwidth, and processing algorithms were not ready to deal with audio and video. With the increasing availability and popularity of CDs, people became excited about creating documents that could include not only text, but also images, audio, and even video**.** That decade saw explosive growth in all aspects of hardware and software technology related to multimedia computing and communication. In the early 1990s, PC manufacturers labeled their high-end units containing some advanced graphics *multimedia* computers. That trend disappeared a few years later because every new computer was a multimedia computer.

Research and development in multimedia-related areas has been around for much longer. Research in speech processing, speech compression, and speech recognition was fueled first by telephony and then by digital sound applications. Image and video processing and compression have also been active research and development areas due to digital photos and then digital video.

Before 1990, much of the research in audio and video compression, storage, and communication was driven by broadcast and consumer electronics related to entertainment applications. Inthe 1990s, the idea of combining these sources in a computing environment emerged as a clear possibility. As a result, research in all areas of audio and video received significantly greater emphasis.

When people think of multimedia computing, they usually think of video in a computing environment. This is a narrow perspective on multimedia. Visual information definitely dominates human activities because of the powerful visual machinery that we are equipped with. But, humans use all five senses effectively, opportunistically, and judiciously. A true multimedia system should be able to effectively utilize signals from multifarious sensors and present to users only the relevant data in the appropriate media.

This book takes an *integrative systems* approach to multimedia. Integrated multimedia systems receive input from different sensory, symbolic, and human **//meaning?//** sources in different forms and representations. Users access this information in experiential environments. Early techniques dealt with individual medium more effectively than with integrated media and focused on developing effective techniques for separate individual medium. **//for example, …?//** As the field matures, we are seeing increasing attention on issues that span multimedia. As we will discuss in more detail later, most of the difficult semantic issues become easier to solve when considering integrated multimedia rather than separate individual medium.

# Defining Multimedia

One obvious question that comes to mind at this stage is: Are there some fundamental issues in multimedia computing and communication systems that will provide this integrative perspective? For exploring this, let us consider the problem a bit more precisely.

Consider a system equipped with multiple sensors working in a physical environment. The system continuously receives information about the environment from multiple sensors and uses this information to achieve its goals.

Assume that *S*1, … *Sn* are synchronized data streams from sensors. These data streams have *K* types of data in the form of image sequence, audio stream, motion detector, and so on. Further, let *M*1, …, *Mn* be metadata, including annotations, for each stream. This metadata might include the sensor’s location and type, viewpoint, angles, camera calibration parameters, or any other similar parameters relevant to the data stream. In most cases, feature detectors must be applied to each data stream to obtain features that are relevant in a given application. Let us represent feature stream *Fij*, where *Fij* is the *j*th feature stream from *Si*.

Multimedia computing and communication techniques combine the dataset *Si* and **//its feature stream//** *Fij* using the metadata *Mi* to extract information about the environment required to solve a given problem. In this process, **//the system//** must often combine partial, sometimes uncertain, information from multiple sources to get more complete and reliable information about the environment.

**A defining difference in multimedia from single medium understanding fields like computer vision or audio processing is that partial information from multiple media is correlated and combined to get complete information about the environment. A common experience that most people have is deciding about a thunder and explosion—appearance of a bright light followed by a strong sound is used to detect it. Without correlating the sound with the noise, one can not conclude that there is an explosion or a thunder.**

As we will see, context captured using **//the senses?//** plays a key role in multimedia analysis. And the context can come from some data collection parameters or from other sensory data. In fact, the distinction between content and context is rarely clear—it really depends on the context **//clarify? distinguishing context depends on context?//**.

# Organization of the Book

We organized this book to present a unified perspective on different media sources for addressing emerging applications. The book consists of seven sections, each containing several chapters. We provide pointers to the latest literature, but our main goal here is to present concepts, techniques, and applications that will be useful in building integrated multimedia systems. We believe that the holistic viewpoint presented in this book is essential for understanding, using, and communicating emerging applications that use heterogeneous data from multifarious sources.

## Defining Multimedia Systems

The current stage of the multimedia field brings to mind the parable about the six blind men and the elephant; we therefore define a multimedia system and discuss its main elements. This will help us discuss all elements concretely without losing the whole-system perspective.

## Nature of Information

**Like humans, multimedia systems** gain information and experience through a variety of sensory and other sources. Understanding the relationships among data, information, knowledge, insight, and experience is crucial to being able to use these sources judiciously. We discuss basic elements of information and data source types, including text, audio, images, and video, in the context of multimedia systems. These areas are well established and many other sources provide details on every aspect of representation and processing.Our goal here will be to present the essential elements from those areas and direct readers to sources for more information.

## Creation and Presentation

Although **//multimedia systems? they? (i.e., users)//** acquire original data through sensors and other sources, users often use production environments to edit and create multimedia presentations. Many editing and production environments exist for creating these presentations. Most of these environments were **//originally//** offline, but with the Web and technology advances, users increasingly require runtime environments for creation and presentation.

## Communication

Sensors are often located at geographical locations outside of the processing environment. Users, too, are typically at different geographic locations from the processor. Thus, increasingly, a system’s input, processing, and output elements are at different locations. A large volume of data must therefore be communicated to different locations, making data-compression techniques essential. Fortunately, data compression is an active research area and most of these techniques have responded well to multimedia systems’ needs.

In addition to compression techniques, several networking issues merit mention because different types of media have different requirements due to compression and representation **//requirements//**. All of these issues present challenges to middleware design in network environments.

## Organization and Access

Multimedia systems require a large amount of storage. Data’s distributed nature and presentation issues make storage techniques and architectures serious concerns.

Organizing multimedia data for search and navigation has been a challenge. Even organizing individual components such as audio, images, and video presents challenges. In the last few years, researchers have begun focusing on spatiotemporal multimedia data. Systems that handle this data will require different access environments and different navigation and presentation mechanisms from those used in current databases and search engines.

## Systems Issues

When designing multimedia systems, several interesting architectural and systems-related challenges arise. In many applications, users want personalized presentations that might involve large volumes of data. This usually presents interesting challenges related to storage, processing, and communication architecture. Further, at presentation time, all data must be synchronized.

Digital rights management is also becoming a major issue. If the rights of content owners are not protected, all the technology will remain unused in entertainment and some other application areas. Similarly, businesses and other organizations will not use these systems if privacy and security issues are not resolved.

## Application Systems

After discussing different elements of multimedia systems, we present a few applications that use multimedia systems. These applications are demonstrative in nature and are selected to show how different elements are put together.